



NATIONAL
CENTER *for* ANALYSIS of LONGITUDINAL DATA *in* EDUCATION RESEARCH

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

A program of research by the American Institutes for Research with Duke University, Northwestern University, Stanford University, University of Missouri-Columbia, University of Texas at Dallas, and University of Washington



*Learning that Lasts:
Unpacking Variation in
Teachers' Effects on
Students' Long-Term
Knowledge*

BENJAMIN MASTER,
SUSANNA LOEB,
JAMES WYCKOFF

Learning that Lasts: Unpacking Variation in Teachers' Effects on Students' Long-Term Knowledge

Benjamin Master
Stanford University

Susanna Loeb
Stanford University

James Wyckoff
University of Virginia

Contents

Acknowledgements.....	ii
Abstract.....	iii
Introduction.....	1
Hypotheses and Research Questions	7
Data.....	10
Methods	14
Results.....	20
Conclusions and Discussion.....	30
References.....	34
Tables.....	36
Appendix Tables	43

Acknowledgements

This research was supported by the National Center for Analysis of Longitudinal Data in Education Research (CALDER) funded through Grant R305C120008 to the American Institutes for Research from the Institute of Education Sciences, U.S. Department of Education.

CALDER working papers have not undergone final formal review and should not be cited or distributed without permission from the authors. They are intended to encourage discussion and suggestions for revision before final publication.

The views expressed are those of the authors and should not be attributed to the American Institutes for Research, its trustees, or any of the funders or supporting organizations mentioned herein. Any errors are attributable to the authors.

CALDER • American Institutes for Research
1000 Thomas Jefferson Street N.W., Washington, D.C. 20007
202-403-5796 • www.caldercenter.org

Learning that Lasts: Unpacking Variation in Teachers' Effects on Students' Long-Term Knowledge

Benjamin Master, Susanna Loeb and James Wyckoff

CALDER Working Paper No. 104

January 2014

Abstract

Measures of teachers' "value added" to student achievement play an increasingly central role in k-12 teacher policy and practice, in part because they have been shown to predict teachers' long-term impacts on students' life outcomes. However, little research has examined variation in the long-term effects of teachers with similar value-added performance. In this study, we investigate variation in the persistence of teachers' value-added effects on student achievement in New York City. We separate persistent effects into general effects that improve both the subject taught (math or English language arts (ELA)) and the other area of measured achievement and subject-specific effects which improve only the subject taught. Two findings emerge. First, a teacher's value-added to ELA achievement has substantial crossover effects on long-term math performance. That is, having a better ELA teacher affects both math and ELA performance in a future year. Conversely, math teachers have only minimal long-term effects on ELA performance; their effects are far more subject-specific. Second, we identify substantial heterogeneity in the persistence of English Language Arts (ELA) teachers' effects across observable student, teacher, and school characteristics. In particular, teachers in schools serving more poor, minority, and previously low-scoring students have less persistence than other teachers with the same value-added scores. Moreover, ELA teachers with stronger academic backgrounds have more persistent effects on student achievement, as do schools staffed with a higher proportion of such teachers. The results indicate that teachers' effects on students' long-term skills can vary as a function of instructional content and quality in ways that are not fully captured by value-added measures of teacher effectiveness.

Introduction

Teachers play an important role in students' academic achievement gains (Rivkin, Hanushek, and Kain, 2005; Goldhaber, 2002). Most estimates suggest that assignment to a teacher whose "value-added" performance is one standard deviation above the average teacher will raise student achievement by 0.1 to 0.2 of a standard deviation in a single school year (Kane and Staiger, 2008; Rockoff, 2004; Nye, Konstantopoulos, and Hedges, 2004). In addition, recent work by Chetty, Friedman, and Rockoff (2011) shows that teacher quality as measured by teachers' value added to student achievement predicts teachers' impacts on students' long-term life outcomes, such as college attendance and quality, lifetime income, and the likelihood of becoming a teen parent. For instance, they find that a one standard deviation improvement in the value-added score of the teacher that a student receives in middle school corresponds to a one percent increase in their expected earnings by age 28.

In light of the importance of teacher quality, and of the correlation between teachers' effects on tested achievement and students' long-term outcomes, policy makers and district leaders are increasingly utilizing teacher value-added measures to evaluate differences in teacher performance. Most notably, both competitive grants provided by the federal Race to the Top Initiative and recent waivers to federal No Child Left Behind accountability standard are spurring reforms to teacher evaluation practices in schools, and encouraging the use of value-added measures to assess teachers' quality. These and other reforms by states and local districts reflect a growing reliance on value-added measures as a central performance benchmark in systems designed to gauge and respond to the performance of k-12 teachers.

While teachers' value added to tested student achievement predicts their long-term impacts on students, it is unlikely that effects on single-year achievement gains, which is what value-added scores

measure, are the only mechanism by which teachers help students to succeed in the long run. Indeed, most of a teacher's immediate effect on test scores does not persist, but fades out within a few years (Chetty et. al, 2011; Jacob, Lefgren, and Sims, 2010; Rothstein, 2010). Instead, teachers' ability to impart tested content and skills to students likely only partially aligns with their impact on outcomes that are not immediately assessed, including students' longer term academic achievement. Just as there is variation in teachers short-run effects, we would expect to see heterogeneity in the long-term instructional impacts of teachers, even among those with similar value added to short term student achievement. This variation in long-term effect for teachers with similar value-added scores could occur if, for example, two teachers prepare students comparably in terms of tested material, but one teacher also imparts additional knowledge or skills that are useful but not immediately assessed. In this paper, we examine this variation in persistence of teachers across subject area and across teacher, student, and school characteristics.

An understanding of the extent to which value-added scores capture variation in teachers' long-term impacts on students can shed light on the validity of using value-added approaches to measure teacher effectiveness. If there is substantial and systematic variation in the long-term effects of teachers with comparable value-added scores, then policies and management systems that rely heavily on value-added measures could benefit from additional measures capturing longer-term effects. In addition, emerging evaluation and accountability systems may incentive educators to focus excessively on short-term tested outcomes in ways that are not ultimately beneficial for students. Monitoring the alignment between impacts on short and longer term outcomes can help to ensure that the measures employed to assess teacher and school effectiveness are aligned with ultimate instructional priorities and values.

Heterogeneity in the Relationship Between Value Added and Teachers' Long-term Effects

A small but growing body of research has shown that the relationship between teachers' value added to short-term achievement gains and their long-term impacts on students can vary systematically as a function of the type and quality of knowledge that students gain. In particular, in the one study that links teacher value-added measures to students' long-term life outcomes, Chetty et. al (2011) find that teachers' value-added effects in English language arts (ELA) predict substantially greater impacts on long-term student outcomes than math teachers' value-added effects. That is, an English teacher who raises students' test scores by 1 standard deviation has an impact on long-term life outcomes approximately 1.7 times that of a math teacher who does the same. Because the variance of English teacher effects is lower than that of math teachers, they find that the effect of having a teacher who is 1 standard deviation above the mean in terms of value-added quality is similar across subjects. Nevertheless, by-subject heterogeneity in the import of teachers' effects suggests that the long-term benefits that students gain from instruction may vary meaningfully depending on the type of knowledge that they acquire. Improvement in ELA skills may yield larger benefits for students due to the broad relevance of reading skills across diverse subject areas, as well as in non-academic contexts. Abedi and Lord (2001) find that adjusting the linguistic complexity of mathematics achievement tests influences students' performance, with larger impacts on both English language learner students and students of lower socio-economic status.

In addition to heterogeneity in the import of teachers' effects by subject, Chetty et al (2011) find that having a high value-added teacher predicts greater long-term benefits for students of higher socio-economic status (SES) than for lower-SES students. Importantly, they find this systematic difference even though the effect of teachers' on test scores in the year students have the teacher is similar between these groups. Students of different socio-economic status, then, benefit differently from the same observed test-score gain. However, the mechanisms underlying this result are unclear. Differences in long-term effects could stem from higher SES students' differential ability at retaining or

making profitable use of tested skills. Alternately, this heterogeneity could be driven by unobserved differences in teachers' instructional quality that correlate with students' SES, but are not captured by measured effects on short term achievement.

Teachers' instructional quality may be a meaningful driver of variation in the relationship between their short-term and long-term effects on student achievement. Corcoran, Jennings, and Beveridge (2011), for example, find that, in a context in which accountability pressures are salient, value-added measures that are based on high-stakes tests provide less-accurate information regarding teachers' long-term effects than do value-added measures of the same teachers that are based on low-stakes tests. This result provides evidence that incentives to produce short-term achievement gains can lead some teachers to detrimentally modify their instruction to focus on short-term tested knowledge at the expense of longer-term student learning. A more extreme example of this trend is apparent in the absence of long-term impacts for teachers who cheat in order to generate high short-term value-added results (Jacob and Levitt, 2003; Chetty et al., 2011). Separately, Carrel and West (2010) find - in a university context - that veteran professors tend to have lower short-term value added, but higher long-term value added, relative to their less experienced peers, perhaps stemming from greater focus on long-term skill development among experienced professors. Collectively, this body of evidence demonstrates that teachers' instructional practices can influence their short-term value-added performance in ways that do not correspond with long-term success for students.

Persistence of Teacher Value-Added Effects

One way to explore the relationship between teachers' impacts on short-term and long-term student outcomes is to examine variation in the persistence of value-added effects on academic achievement after a student leaves a teacher's classroom. To investigate persistence, we borrow a conceptual and methodological framework described in Jacob, Lefgren, and Sims (2010). In their framework, persistent academic effects can be understood as teachers' development of students' *long-*

term knowledge and skills that are relevant to both contemporaneous and future achievement tests.

Short-term knowledge gains, in contrast, decay completely and have no effect on future tested achievement, even in the following year. This short-term knowledge may include facts such as state capitals that may have some long-term benefits to students but do not contribute to (or serve as a base for) future academic achievement or test-taking skills that are specific to the test in question and probably have no long-term benefits for students.

Effects on long-term knowledge may be categorized further into *subject-specific* content learning that is assessed repeatedly across exams within the same subject area and more *general* learning that raises a student's underlying ability or motivation and improves their achievement across a broader range of material. Some teachers may be particularly effective at imparting subject-specific content knowledge that improves students' long-run performance in that subject. For example, a clear understanding of fractions can help students in algebra and geometry but may not have a strong impact on students' writing or critical reading skills. Other teachers may be less successful at these subject-specific topics but may help to develop students reasoning skills or motivation for learning that have benefits across subjects in future years.

Variation in teachers' impacts on long-term knowledge – both subject-specific and general – is important for two reasons. First, if some teachers have more persistent effects on achievement than others, they will have larger cumulative impacts on students' academic success than their same-year value-added performance implies. By not accounting for differences in the persistence of teachers' effects in high-stakes teacher evaluations, schools may unconstructively incentivize teachers to focus only on short-term results, and may also make less efficient personnel decisions on the basis of those short-term results. Second, the persistence of teachers' effects on achievement may be a useful indicator for other long-term effects that teachers have on students' life outcomes. That is, teachers whose effects persist more in the area of academic achievement may also have more persistent effects

on students in general, particularly if their effects on long-term knowledge represent learning that is applicable across a broad range of contexts. If this is the case, then studying variation in the persistence of teachers' effects on academic achievement can provide insights into some of the mechanisms by which teacher quality translates into long-term student success.

Prior research on the persistence of k-12 teachers' value-added effects indicates that most of the effects do not persist into future school years, and that teachers' long-term effects are not necessarily aligned with their initial value-added performance. Across a range of samples and methodologies, most estimates of persistence are that only approximately one-third to one-fifth of elementary or middle school teachers' value added persists into the subsequent school year, with lower rates of decay after that (McCaffrey et al, 2004; Rothstein, 2010; Jacob et. al, 2010; Konstantopoulos 2011; Kinsler, 2012).¹ Rates of persistence of value-added effects appear to be comparable within the subject areas of Reading, Math, and Science, though no study that we're aware of has examined the persistence of general knowledge in comparison to subject-specific knowledge across subject areas. Separately, Rothstein (2010) investigates the correlation between teachers' initial value-added effects on achievement and their longer term value-added effects over two or three years, and finds that initial and long-term effects are only modestly aligned, with sampling-error-adjusted correlations between 0.3 and 0.5.² This result suggests that there may be substantial variability in teachers' long-term academic impacts that is not reflected in their short-term value-added performance.

Variation in the persistence of teachers' effects on student academic achievement may arise through a variety of mechanisms. First, we would expect persistence to vary according to the degree of overlap in content between one test administration and the next, and that some types of long term

¹ There is, however, some variation in estimated persistence in the literature. Using a mix of experimental and non-experimental data, Kane and Staiger (2008) estimate persistence of value added effects that are close to 50% after one year, while Lockwood, et al. (2007) estimate persistence parameters that are less than 1/5th.

² Rothstein (2010) estimates teachers' effects β in a single year and classroom in comparison to their cumulative effects on students over multiple years, assuming a constant rate of decay λ for teachers' effects. For example, he estimates the cumulative effect of a third grade teacher on fifth grade students as $(\beta_{33_c}\lambda + \beta_{34_c})\lambda + \beta_{35_c}$.

knowledge may be more robust to differences in test content. Second, teachers' instructional quality or curriculum may vary, with some teachers placing more emphasis on increasing students' long-term knowledge that is more likely to persist, while others focus more on short-term tested knowledge that is less relevant to future performance. An oft-cited example of such quality variation would be the distinction between a hypothetical teacher who "teaches to the test," with a focus primarily on test-taking strategies and superficial content, versus one who focuses on instilling deeper understandings and an affinity for learning in students. A third factor that may drive variation in the persistence of learning gains associated with teachers' value-added scores is the rate at which different students forget long-term knowledge, regardless of the quality of instruction that they initially received. Such differential forgetfulness could result from students' innate abilities or from instructional or other contexts that support different rates of knowledge retention. Finally, students may also differ in their ability to acquire long-term knowledge, even when presented with identical instruction. For example, while a teacher might provide a similar instructional experience to all her students, some students may tend to acquire a more superficial grasp of the material, while others may be more receptive to opportunities for deeper learning.

Hypotheses and Research Questions

In this study, we expand upon the limited extant research about the heterogeneity of teachers' long-term effects by examining variation in the persistence of student learning associated with teacher value added across the New York City (NYC) public school system. We start by replicating earlier studies of average teacher persistence in order to place our sample in the context of earlier studies. We then explore three hypotheses concerning variation in persistence suggested by both theory and the prior literature. First, teachers' long-term effects may vary in the extent to which they are subject-specific in comparison to more general knowledge that affects multiple subjects. In light of evidence that effects

on students' ELA knowledge have greater long-term benefits for students than effects on mathematics knowledge (Chetty et al., 2011), we hypothesize that ELA teacher effects may consist of a greater portion of knowledge that is relevant across subject areas. If ELA crossover is greater, then we would expect ELA teachers to influence students' future performance in math to a greater extent than math teachers influence future student performance in ELA.

Second, we anticipate that differences in curriculum or instructional quality may predict the mix of short-term versus longer-term knowledge that students learn. Although we lack detailed measures of instructional practice across NYC schools, we are able to partially test this hypothesis by leveraging administrative data. First we can look at variation across teachers with different characteristics. For example, teachers vary in the types of colleges they attended, as well as their performance on academic and licensure tests. These characteristics serve as distal measures of the foundational knowledge and ability that teachers bring to bear in their work. Prior research has shown mixed and weak results when predicting teachers' short-term value added with measures of teachers' academic ability (e.g. Goldhaber, 2008, Buddin and Zamarro, 2009, Harris and Sass, 2011). Nevertheless, these characteristics may predict a greater focus on long-term knowledge and thus higher value-added persistence. In addition, we can look at variation across schools with different characteristics. Because school-wide policies and norms may influence the curriculum or instruction that students receive, it is possible that observable characteristics of schools will predict differences in value-added persistence, distinct from variation at the individual teacher level. This school-level differentiation could result, for instance, if some schools encourage more short-term "teaching to the test," while others emphasize a curriculum focused more on long-term knowledge.

Finally, the persistence of value-added scores may vary as a function of students' characteristics such as their socio-economic status or their prior academic ability. Students with fewer supports at home, for example, may forget more of the knowledge they gained from their teacher in school. Extant

literature exploring this hypothesis is limited and shows mixed results. Jacob et. al (2010) do not identify meaningful variation in persistence as a function of students' race or free-lunch status in North Carolina. However, Chetty et al. (2011) find that lower-SES students benefit less in the long-term from teachers' effects on their academic achievement. Returns to short-term achievement gains due to teachers' value-added for low-SES and lower-achieving students may be lower as a result of lower quality instruction. For instance, lower achieving, poor, or minority students might more often be assigned to teachers or schools that prioritize short term achievement over long-term learning. This teacher-driven or school-driven explanation could be the case, for instance, if their teachers or schools experience greater accountability pressure to attend to short-term achievement outcomes. Alternately, it is possible that students could experience less persistent academic benefits from instruction due to student-level factors. Low-income or low-achieving students may retain long-term knowledge at lower rates than other students, or they may be less skilled at acquiring long-term knowledge when presented with the opportunity.

In order to test our hypotheses, we consider the following research questions:

- 1) What is the persistence of teachers' value-added effects in NYC, both within the subject areas of English language arts and math, and across subject areas?
- 2) To what extent does value-added persistence vary as a function of teachers' ability, as measured by their teacher-licensure exam scores, undergraduate college competitiveness, and scholastic achievement test (SAT) scores?
- 3) To what extent does value-added persistence vary as a function of students' socio-economic status or their prior academic achievement levels?
- 4) To what extent does variation in value-added persistence stem from students' differential rates of forgetting previously acquired long-term knowledge?

- 5) To what extent does variation in teachers' value-added persistence stem from school-level characteristics, including the socio-economic and prior ability characteristics of students who attend a school or the average academic ability of teachers employed at a school?

We find that teacher value-added effects in ELA have substantial cross-over effects on long-term student math performance, but that the reverse is not true for math teachers' effects on long-term ELA performance. ELA teachers then influence not only subject-specific content, but also general long-term skills that affect future achievement across subjects. In addition, we identify substantial heterogeneity in the persistence of ELA teachers' effects across observable student, teacher, and school characteristics. In particular, teachers in schools serving more poor, minority, and previously low-scoring students have less persistence than other teachers with the same value-added scores. Moreover, ELA teachers with stronger academic backgrounds have more persistent effects on student achievement, as do schools staffed with a higher proportion of such teachers. While we cannot definitively distinguish between teacher, student, and school-level factors that may drive variation in value-added persistence, our evidence suggests that differences in instructional quality play a critical role, particularly at the school level. Students vary in the rate at which they forget long-term knowledge, but this variation is modest, and is unlikely to be a major driver of the differences that we observe in their teachers' value-added persistence. Overall, our results demonstrate that teachers' effects on students' long-term skills can vary substantially and systematically, in ways that are not fully captured by short-term value-added measures of instructional quality.

Data

Administrative Data

In order to investigate variation in the persistence of teachers' value-added effects, we draw upon extensive administrative data about students, teachers, classrooms, and schools from the New York City

Department of Education (NYCDOE) and the New York State Education Department (NYSED). Our primary data set includes approximately 700,000 students in third through eighth grade in the NYC school system from school years (SY) 2003-04 through SY 2011-12. Our sample includes data on these students' annual standardized achievement test scores in ELA and math. The district data also identifies students' primary teacher and classroom in each year and subject area from SY 2004-05 onward. For the purposes of our analysis, we standardize students' achievement test scores within each grade, subject, and year.

The content assessed by the New York state ELA and math achievement tests has been aligned with the state's content standards in grades three through eight, and has included a considerable portion of questions that require open responses by students, particularly in math. The ELA exams primarily have assessed students' comprehension of reading passages and writing ability, while math exams have addressed a range of topics including number sense, algebra, probability, and geometry, with overlapping topics across grades.³ Depending on the grade level of the assessment, anywhere from 13 to 33 percent of ELA exam points have come from short or extended response items that required written responses, rather than multiple choice answers. In math, short and extended response problems that require students to show and/or explain their work have represented anywhere from 40 to 60 percent of students' scores on the exams.⁴

We are able to match students and teachers to additional demographic, behavioral, and personnel data that we utilize in our analyses. For students, these include their race, ethnicity, and home language, as well as their absences, suspensions, school transfers, free or reduced price lunch status, disability or special education status, and English language learner status in each school year. For teachers, we have

³ From SY 2004-05 through SY 2009-10, math and ELA exams were administered annually in March and January, respectively. From SY 2010-11 onward, both exams were administered in April. As a specification check, we examined persistence separately across these two time periods, but our findings were similar in each.

⁴ In both subject areas, exams are staggered such that tests are longer and include more open-ended questions (and in ELA, more writing prompts) in non-adjacent grades four, six, and eight.

access to unusually rich personnel data, including information about their academic ability. This data includes Barron’s rankings of the relative competitiveness of teachers’ undergraduate college, and teachers’ performance on New York City’s Liberal Arts and Sciences (LAST) licensure exam, which assesses individuals’ knowledge and skills in five areas ranging from science and math to written analysis and expression. For approximately 45 percent of teachers, we also have data on their SAT I ability scores in Verbal and Math.

We estimate teacher value-added effects using our full sample of student-year records in grades four through eight for whom current and prior year achievement data is available. However, because we are investigating the persistence of teachers’ value-added effects on student achievement in the year *after* they teach a student, the sub-sample for most of our analyses consists of student-year observations in which we can identify both current and prior year student achievement data, as well as a prior-year teacher with an appropriate value-added score, as detailed in the Methods section below. In practice, these requirements reduce our analysis of teachers’ persistent effects on student outcomes to approximately one million student-year observations across grades five through eight, from SY 2005-06 through SY 2011-12.

Descriptives

Table 1 provides an overview of student, teacher, and school characteristics from our analytical sample. Consistent with overall student demographics in the NYC public school system, our sample includes a majority (71.8 percent) of students whose limited family income qualifies them to receive free or reduced price school lunches. The majority of students are either black (30.2 percent) or Hispanic (36.3 percent). The average teacher in our sample has more than 7 years of experience in the district. Approximately 30 percent of teachers attended an undergraduate institution ranked as either “most competitive” or “competitive,” while the other 70 percent attended institutions ranked as either “less

competitive” or “not competitive.” Average SAT Verbal and Math scores in the district are slightly below the national average.

In order to provide additional context about the distribution of teachers and students in the district, we present in Table 2 the correlations between our observable teacher characteristics and the characteristics of their classrooms and schools. By design, our teacher value-added measures are largely uncorrelated with observable measures of the socio-economic or prior achievement makeup of their own classrooms or schools. In addition, our measures of teacher’s academic ability, such as their undergraduate competitiveness or licensure test scores, are also uncorrelated with measured teacher value added. Teachers’ academic ability measures are slightly negatively correlated with their students’ socio-economic or prior ability characteristics; while, the school-wide average of teachers’ undergraduate competitiveness is more substantially negatively correlated with the school-wide percent of students whose twice-lagged achievement scores are below the mean, at -0.214. This correlation indicates that there is some overlap between those schools that serve more low ability students and schools that do not recruit many teachers from highly competitive undergraduate institutions.

Accountability Context

This study is descriptive in nature and we do not test whether specific causal factors affect value-added persistence. Nevertheless, because accountability pressures may influence persistence (Corcoran et al., 2011), we include here some context on the district’s accountability practices during this period. Like other districts, NYC has had to respond to accountability measures included in the NCLB act of 2001. In addition, beginning in SY 2007-08, the district began implementing its own school-level accountability system in which schools receive formal Progress Reports and are evaluated via an A-F grading system. School grades are based on a combination of student achievement levels and growth, as well as school environmental factors (New York City Department of Education, 2013). Schools that receive low grades face high stakes consequences, including potential closure, leadership changes, and options for parents

to switch schools.⁵ Research on the effects of this accountability system suggests that receiving low accountability grades has spurred increases in NYC schools' student achievement (Rockoff and Turner, 2008).

Methods

Teacher Value-Added Measures

In order to examine variation in the persistence of teachers' effects, we first generate teacher value-added measures to estimate the effects of each teacher on tested student achievement in each year. We intentionally employ a value-added model that is similar to that used by the NYC Department of Education to evaluate their teachers' performance (University of Wisconsin, 2010). Conceptually, this model compares teachers to other "similarly circumstanced" teachers by first predicting students' achievement with both prior achievement measures and a range of observable student, classroom, and school characteristics that may influence their achievement, and then attributing the remaining unexplained variation in student performance to individual teachers.

We compute value-added scores in three stages. In the first stage we estimate the coefficients λ for students' pretests and β for student-level characteristics on students' posttest scores. To estimate these coefficients, we regress posttest Y_t of student i in classroom c with teacher j in school s at time t on their same-subject pretest Y_{t-1} , other-subject pretest Y_{t-1}^{alt} , a vector of student-level time varying and time invariant variables X , and a set of indicator variables representing individual classroom fixed effects π , which can be expressed as:

$$Y_{icjst} = \lambda Y_{it-1} + \lambda^{alt} Y_{it-1}^{alt} + \beta X_{it} + \pi_{cjs} + \varepsilon_{icjst} \quad (1)$$

⁵ For around 200 "high-needs" schools, NYC also implemented a school-level performance pay incentive system between SY 2007-08 and SY 2009-10. This system rewarded schools' staff with bonuses on the basis of student achievement. However, this program did not meaningfully influence teacher practices or school achievement levels (Marsh, Springer, McCaffrey, Yuan, and Epstein, 2011).

Our student-level characteristics include students' gender, race, an indicator for whether the student's home language is English, student eligibility for free or for reduced price lunch, student disability status, English language learner status, an indicator for whether the student switched schools in the prior year, and the number of prior-year absences for the student. Because the effects of characteristics may vary across grade levels, we also include interactions of each student characteristic with each individual grade level.

In the second stage, we use the estimated coefficients λ and β from our first stage to compute a new left-hand side variable q_{icjst} , where

$$q_{icjst} = Y_{icjst} - \lambda Y_{it-1} - \lambda^{alt} Y_{it-1}^{alt} - \beta X_{icjst} \cdot q_{icjst}$$

is, then, the difference between the student's actual score and what we would predict it to be given background characteristics and prior performance. We then regress q_{icjst} on a vector C of classroom-level characteristics, time-varying school-level characteristics K , and individual year and grade dummy indicators:

$$q_{icjst} = \gamma C_{cjst} + \eta K_{st} + \alpha_t + \rho_g + w_{icjst} \quad (2)$$

Classroom-level characteristics include the racial and home language composition of the classroom, class size, the percent of students who are free or reduced price lunch eligible, percent of students who are English language learners, the class average number of prior year absences, the class average prior year test scores in the same and alternate subject, and the standard deviation of classroom test scores in each subject. As we did for the student covariates, we include interactions of each classroom characteristic with each grade level indicator. School characteristics include total enrollment, the percent of black, white, and Hispanic students in the school, and a control for the percent of students eligible for free or reduced price lunch. When running this regression, we specify a classroom random effect to take into account that errors are correlated within classrooms. From this regression, we obtain

an estimate of w_{icjst} , that represents the residual test score variation for each student in each year that is not explained by our observable student, classroom, or school characteristics.

In our third stage, we estimate individual teacher value-added measures in each year, τ_{jt} , by attributing all remaining variation in students' post-test scores to a combination of the individual teacher effects and error. This can be expressed as:

$$w_{icjst} = \tau_{jt} + \varepsilon_{icjst} \quad (3)$$

We obtain estimates of the error term ε_{icjst} by subtracting each teachers' mean effect, τ_{jt} , from the estimates of w_{icjst} .⁶ Finally, we standardize our teacher-by-year effect estimates across our sample to have a mean of zero and a standard deviation of one.⁷ We include in our analysis only teacher-by-year effects that are based on at least 5 students.

Alternate Value-Added Model Specifications

There is some evidence that the choice of a specific value-added model is less important than the choice of control variables (see for example, Goldhaber and Theobald, 2012). Nonetheless our value-added model specification may influence our resulting estimates of teacher value-added persistence and heterogeneity in persistence. To account for this possibility, we test the robustness of our findings to two alternative model specifications. First, we consider an alternative model that includes school fixed effects (in a separate stage) in lieu of school characteristic controls. This approach allows us to examine persistence in teacher effects that are distinct from individual school effects. Second, we consider an alternative model that predicts student posttests purely as a function of pretest scores and year, grade, and teacher fixed effects, with no additional control variables. This model allows us to compare results for our baseline model that evaluates "similarly circumstanced" teachers against results for a model that

⁶ The standard errors computed under this approach ignore error that comes from having used estimates of λ , β , C , and K to control for pre-tests, student-level variables, classroom-level variables, and school-level variables rather than the true values. However, given our instrumental variables approach to estimating persistence, we are less concerned with error in our value added measures.

⁷ Prior to standardizing, the standard deviation of our ELA teacher-by-year value added measures is 0.23, while the standard deviation for Math teachers' value added measures is 0.27.

instead attributes all of a student's achievement gains to individual teachers' effects, regardless of their context or the students they serve.

Estimating the Persistence of Teacher Value-Added Effects

We estimate the persistence of teachers' value-added using an instrumental variables approach described by Jacob et. al (2010). As previously discussed, these authors conceptualize students' tested knowledge as a combination of "short term" knowledge that has no observed impact on future achievement, and "long-term" knowledge that is relevant to both contemporaneous and future achievement tests. In their formulation, observed student achievement Y in a given period t represents a combination of that student's long-term knowledge from a prior period and all contemporaneous impacts (including teachers' effects) that influence both their long and short term knowledge in the current period:

$$Y_t = \theta y_{l,t-1} + \eta_t^l + \eta_t^s \quad (4)$$

Here, current achievement is a function of contemporaneous impacts η_t^l and η_t^s on long and short term knowledge, as well as long-term knowledge in the prior period $y_{l,t-1}$, which carries forward with some rate of decay $(1-\theta)$.

In practice, we do not directly observe long-term knowledge, but rather the sum of long-term and short-term knowledge assessed in the prior period, Y_{t-1} . In light of this, the authors describe how an ordinary least squares (OLS) coefficient θ_{OLS} for a regression of current achievement on prior achievement converges to the following:

$$\text{plim}(\hat{\theta}_{OLS}) = \theta \left(\frac{\sigma_{yl}^2}{\sigma_{yl}^2 + \sigma_{ys}^2} \right) \quad (5)$$

This equation shows that because prior knowledge consists of a mix of long- and short-term knowledge, the OLS coefficient will be attenuated to the extent that Y_{t-1} consists of short-term, rather than long-term knowledge. In lieu of an OLS estimate of the persistence of observed knowledge, Jacob et. al

(2010) use an instrumental variables approach to estimate the decay of prior long-term knowledge, using twice lagged achievement Y_{t-2} as an instrument for Y_{t-1} . This estimator, which we refer to as $\hat{\theta}_{LT}$, purges Y_{t-1} of its short-term knowledge component. Jacob et. al (2010) estimate (and we also find) that almost all of a student's long-term knowledge persists between one year and the next, with a value of $\hat{\theta}_{LT}$ close to 1. This serves as a benchmark for our subsequent estimation of teachers' effects on long-term knowledge.

Following a similar approach, we can estimate the proportion of a teacher's effect that consists of long-term knowledge by instrumenting each student's lagged knowledge Y_{t-1} with their lagged teacher's contribution (value-added) to that knowledge. The lagged teacher's total contribution to a student's lagged knowledge is a combination of her contribution to long- and short-term lagged knowledge, expressed as $M_{t-1} = \mu_{t-1}^l + \mu_{t-1}^s$. Thus, the second stage estimator $\hat{\theta}_M$ converges to:

$$\text{plim}(\hat{\theta}_M) = \theta \left(\frac{\sigma_{\mu l}^2}{\sigma_{\mu l}^2 + \sigma_{\mu s}^2} \right) \quad (6)$$

Given an estimate of θ that is close to 1, $\hat{\theta}_M$ approximates the fraction of teacher value-added that is attributable to long-term, rather than short-term, knowledge creation.

In practice, student assignment to teachers is nonrandom, and therefore the measured quality of a student's lagged teacher may be correlated with the quality of their current teacher. To minimize possible bias in our teacher persistence estimates due to nonrandom assignment, we include in our instrumental regression to estimate $\hat{\theta}_M$ additional controls for both student level covariates χ and for contemporaneous classroom fixed effects π (which subsume school, year and grade fixed effects). In addition, because teachers' value-added scores in any given year include estimation error that is correlated with other classroom-specific learning shocks in that year, we calculate, for each student in each period, their lagged teachers' average value-added score across all years *other* than the one in

which they taught that student, expressed as $T_{ijt-1} = \sum_{y \neq t-1} M_{jy}$. The second-stage equation for estimating the persistence of teacher value-added then becomes:

$$Y_{icjt} = \theta_M Y_{it-1} + X_{it} + \pi_{cjt} + \varepsilon_{ijt} \quad (7)$$

Where the values of T_{ijt-1} for the lagged teachers serve as the excluded instruments for prior test scores in the first stage. In this formulation, persistence is a function of variation in the quality of the lagged teacher, distinct from the effects of the student's teacher or school in the current year.⁸

In order to estimate cross-subject persistence, we modify equation 7 by replacing our outcome measure, Y_{icjt} , with a student's achievement in the alternate subject, Y_{icjt}^{alt} . Thus, for example, we model students' current math achievement as a function of their prior-year ELA achievement, instrumented by their lagged ELA teacher's value-added score. In addition, when predicting current math achievement, we include classroom fixed effects corresponding to their current-year math classroom assignment, rather than their ELA classroom. We do the reverse when estimating the persistence of lagged math teachers' value-added on students' current ELA achievement. We follow a similar procedure to estimate the general persistence of long run knowledge across subjects.

Estimating Heterogeneity in Teacher Value-Added Persistence

In order to test for heterogeneity in the persistence of teachers' value-added effects across our teacher and student characteristics of interest, we modify the first stage of our instrumental variables equation by replacing T_{ijt-1} with two interaction terms. The first term is set equal to T_{ijt-1} when the binary teacher or student characteristic is equal to 1, and 0 otherwise, while the other equals T_{ijt-1} when the characteristics is equal to 0, and is 0 otherwise. We similarly replace our lagged achievement measure Y_{it-1} from equation 7 with two interacted terms, following the same logic. For example, when

⁸ Jacob et al. (2010) note, however, that our estimates of persistence may still be biased if schools adjust the instructional inputs (other than classroom assignments) that students receive, as a response to the quality of their lagged teacher. This could occur, for instance, if effective teacher raise students' achievement and this in turn leads schools to provide fewer instructional supports to the student.

investigating persistence across poor and non-poor students, we separately instrument poor students' lagged achievement scores with their lagged teachers' value-added, while also instrumenting non-poor students' achievement scores with their own lagged teachers' value-added. In cases where we are missing data on a student or teacher characteristics of interest for certain observations, we include an additional instrument and lagged achievement measure interacted with an indicator for the missing data. In practice, our approach yields very similar results to the alternative method of estimating persistence separately across in-group and out-group samples, which we run as a specification check. We opt to use interactions terms to facilitate a more succinct presentation of our findings. For each teacher or student characteristic of interest, we conduct F-tests to assess whether teacher value-added persistence coefficients for the in-group and out-group are significantly different from each other.

Results

Value-Added Persistence Within and Across-Subjects

We address our first research question by describing the persistence of teachers' value-added effects in NYC, both within the same subject area, and across subjects. Table 3 shows the persistence of teachers' value-added effects alongside an estimate of long-term knowledge persistence that uses twice-lagged test scores as an instrument on once-lagged achievement to predict current achievement. Our estimates for the persistence of teachers' effects in math and ELA are comparable to those from prior studies. For each subject, we find that approximately one-fifth of a teacher's value added to achievement persists into the subsequent school year. As expected, the persistence of long-run knowledge is much higher and very close to 1 in both subjects, which indicates that almost all of a student's previous long-term knowledge in a subject persists over time and does not decay.

We also compare teachers' persistence estimates to an OLS coefficient from a regression predicting current achievement with prior achievement. The OLS coefficient in math of 0.772 is notably higher

than that of ELA, at 0.645, which suggests that the NYS math exams are more closely aligned in terms of knowledge assessed across grade levels than the ELA exams. The greater year-over-year alignment in math scores does not, however, correspond to a higher rate of persistence of math teachers' value-added effects. One possible explanation for this apparent inconsistency could be that math teachers' effects on long-term knowledge more often reflect overlap in explicitly assessed content across grades, while ELA teachers' effects on long-term knowledge is less content-specific, and therefore less dependent on year-over-year test alignment.

While we observe similar levels of within-subject persistence for ELA and math, we see substantial differences between math and ELA teachers in terms of their cross-subject persistence. The persistence of teachers' value-added in ELA on future math performance is quite high. The cross-subject persistence coefficient of 0.149 for ELA teachers is three quarters of the size of the within-subject ELA persistence coefficient of 0.208. This indicates that ELA teachers' long-term effects reflect not only impacts on students' subject-specific knowledge, but also knowledge and skills that are highly relevant to students' future math performance. In contrast, the persistence of teachers' value-added effects in math on future ELA performance is much smaller, with a coefficient of 0.043. Most of math teachers' value-added effects appear to reflect gains in subject-specific knowledge that do not influence students' future ELA performance.

Cross-subject, instrumented coefficients of long-term knowledge are similar for ELA and math, with coefficients of 0.639 for previous long-term ELA knowledge on math achievement, and 0.615 for previous long-term math knowledge on ELA achievement. Not surprisingly, these results indicate that long-term knowledge or skills in one subject – measured as the instrument of twice-lagged test scores on once-lagged test scores in that subject – is not perfectly correlated with future achievement in an alternate subject. However, the similarity of the benchmark measures of cross-subject long-run knowledge persistence across ELA and math contrasts with the substantial difference in the persistence

of math and ELA *teachers'* value-added effects in an alternate subject. In other words, while students' long-term knowledge and skills appear to be fairly portable across subjects, the same is not necessarily true for teachers' effects on students' long-term knowledge and skills. ELA teachers appear to develop more portable long-term knowledge in students than math teachers do.⁹

Heterogeneity in the Persistence of Teachers' Effects

Next we examine whether the persistence of teachers' effects varies as a function of teachers' ability characteristics or of students' socio-economic or prior ability characteristics.¹⁰ In Table 4 we compare the persistence of teacher value-added in ELA on student achievement in the subsequent school year, across teacher and student characteristics. There is substantial heterogeneity in the persistence of ELA teachers' effects on future achievement. For instance, the within-subject value-added persistence of ELA teachers who attended a more competitive undergraduate institution is significantly and substantially higher than that of teachers who attended a less competitive institution, with corresponding coefficients of 0.274 and 0.177, respectively. These estimates indicate that more than a quarter of teacher value-added effects persist into the next year for teachers from competitive institutions, in comparison to a persistence rate of less than one-fifth for teachers from less competitive institutions.

Differences in persistence are similarly large when comparing teachers whose SAT Verbal exam scores or LAST licensure exam scores are in the top third of the teacher distribution, in comparison to lower-scoring teachers. In both cases, higher scoring teachers show greater persistence. Teachers' SAT Math scores predict a somewhat smaller and non-significant difference in persistence for future ELA achievement. It is notable that our teacher ability characteristics predict large differences in ELA

⁹ In preliminary work, we investigated whether the greater cross-subject persistence of ELA teachers' effects is driven by teachers of English Language Learner (ELL) students in particular. It is not. We found that cross-subject ELA persistence estimates are similar even when excluding ELL students from our analysis.

¹⁰ In preliminary analysis, we also examined whether there are differences in value added persistence across novice and more experienced teachers. Consistent with the result of prior research (Jacob et. al, 2010) we found no significant evidence of heterogeneity as a function of teachers' experience in the district.

teachers' value-added persistence, even though they are not themselves correlated with teachers' short-term value-added effects.

Observable student characteristics related to their socio-economic status or prior ability also predict substantial variation in their ELA teachers' value-added persistence. The persistence of achievement gains coming from having an effective teacher is far lower for students who are eligible for free lunch, are black or Hispanic, or whose twice-lagged ELA achievement scores are below the mean. For example, same-subject ELA value-added persistence for students who are eligible for free lunch is 0.161, in comparison to other students at 0.262. We see similar large differences in persistence when we compare black or Hispanic students to other students or when we compare students with below average twice-lagged achievement scores to students of higher prior performance.¹¹ These students may be receiving ELA instruction that is less focused on long-term knowledge, or they may be less skilled at acquiring or retaining long-term knowledge.

The heterogeneity that we observe in the persistence of ELA teachers' value-added – both by teacher test performance and student background – is apparent across subject areas as well as within ELA. Both teachers' ability characteristics and students' socio-economic and prior ability characteristics predict substantial differences in ELA teachers' value-added persistence on math achievement. For instance, the cross-subject persistence coefficient for ELA teachers who attended more competitive undergraduate institutions is 0.207, in comparison to 0.123 for teachers who attended less competitive institutions. Differences in persistence are similar in magnitude across student characteristics. For example, the cross-subject value-added persistence for teachers of students eligible for free lunch is

¹¹ In preliminary analysis, we examined whether the effects of each student characteristic of interest are independent of the other characteristics, by comparing in-group and out-group persistence in terms of one characteristic across two samples defined by a second student-level characteristic. We found that all of our student characteristics predict significant variation in value added persistence, independent of each other. Using the same approach, we also found that differences in persistence stemming from our teacher and student characteristics are also independent from each other. These results are available upon request.

0.109, in comparison to 0.202 for other students. Variation in the persistence of ELA teachers' effects influences students' longer term ELA and math achievement to a similar extent.

The heterogeneity that we observe in ELA teachers' value-added persistence is robust to alternative value-added model specifications (see Appendix Table 1). As described above, the first alternative is a model that includes school fixed effects in lieu of our base model controls for school-level covariates. The second is a value-added model that includes no controls for student, classroom, or school-level covariates. Our school fixed-effect model yields nearly identical results as our base model in all cases. We see the same heterogeneity in ELA teachers' persistence whether we explicitly rank teachers' value-added effects in comparison to peer teachers at the same school, or simply in comparison to other "similarly circumstanced" teachers across the district. Our second alternative model, which ignores contextual factors and attributes all achievement gains solely to teachers, yields substantially higher persistence estimates overall. However, even in this model specification we continue to see significant differences in value-added persistence as a function of our teacher and student characteristics of interest.¹²

Next we investigate the persistence of math teachers' value-added across the same teacher and student characteristics, and present our results in Table 5. In contrast to the heterogeneity that we observe in ELA, math value-added persistence is for the most part the same across our teacher and student characteristics of interest. Teachers' SAT math scores predict a significant difference in value-added persistence and this heterogeneity is in the same direction as it is for ELA. However, even in this case the absolute difference in value-added persistence is minor, with a coefficient of 0.215 for teachers with high SAT math scores versus 0.173 for lower-scoring teachers. The consistency of teachers' value-added persistence in math may stem from the content of the math achievement tests themselves,

¹² Because overlap in assessment content from one grade to the next can influence the persistence of value-added effects, in additional analysis we also examined persistence across grade levels. We found that persistence rates do vary across grades. However, the heterogeneity that we observe in ELA teachers' value-added persistence is largely consistent across grades. These results are available upon request.

which appear more aligned year-over-year than do the ELA exams. If students' long-term knowledge is more explicitly tested in math, then teachers' measured effects may persist similarly even when instruction focuses more narrowly on tested content.

As previously discussed, math teachers' cross-subject value-added persistence is low, with an average of less than five percent of teachers' value-added effect in math persisting into students' subsequent-year ELA scores. This low level of cross-subject persistence does not vary as a function of observable teacher ability characteristics. However, we do see meaningful differentiation in math teachers' cross-subject persistence as a function of their students' socio-economic and prior ability characteristics. Cross-subject value-added persistence for math teachers of black and Hispanic students is only 0.025, versus 0.090 for students of other races. Similarly, cross-subject persistence for students whose twice-lagged test score was below the mean is only 0.014, while persistence for previously higher-scoring students is 0.066. These results suggest that there may be some systematic differences in the long-term knowledge learned or retained by students of math teachers, even if this differential knowledge is not readily apparent in future math exam scores.

Heterogeneity in the Decay of Students' Long-term Knowledge

The variation that we observe in the association between teachers' value-added effects and students' longer-term achievement may reflect a variety of underlying mechanisms, including differences in the type of the instruction that students receive or differences in students' ability to acquire or retain long-term knowledge. While we cannot definitively distinguish between differences in instructional quality and students' differential skill at acquiring long-term knowledge, we can directly assess the rate at which different students forget previously acquired long-term knowledge. We investigate this potential driver of variation in teachers' long-term effects by examining the persistence of long-term knowledge in math and ELA across our observable student and teacher characteristics. If differences in student forgetfulness were primarily responsible for the substantial variation in value-

added persistence that we observe, then we would expect that some students' would have much lower persistence of long-term knowledge in general, with coefficients substantially less than 1. We present estimates of long-term knowledge persistence for different groups of students in Table 6.

We find that some student characteristics do predict significant differences in the decay rate of long-term knowledge in both math and ELA. Across both subject areas, black and Hispanic students and students with lower prior ability have significantly lower persistence of long-term knowledge. For example, the coefficient on long-term knowledge for previously-lower achieving students is 0.922 in ELA, in comparison to a coefficient of 0.968 for other students. In math, the coefficient on long-term knowledge for previously lower-achieving students is 0.905, versus 0.947 for other students. These results suggest that some portion of students' cumulative academic achievement stems from how effectively they retain knowledge over time. However, the magnitude of the differences in student forgetfulness is small relative to the variation in ELA teachers' value-added persistence that we observe across the same student characteristics, and differential forgetfulness is similar across both math and ELA.

We find no significant differences in long-term knowledge decay across students that have teachers with different ability characteristics, in either math or ELA. Overall, differential retention of long-term knowledge does not appear to be the primary driver of the heterogeneity that we observe in ELA teachers' long-term effects. Moreover, the differences in student forgetfulness that we observe may themselves stem from differences in the effectiveness of some teachers or schools at reinforcing students' prior long-term knowledge gains.

Heterogeneity in Persistence Across School Characteristics

Our results thus far provide evidence of substantial heterogeneity in ELA teachers' value-added persistence, but it is uncertain the extent to which this stems from differences in instructional quality or from unobserved student ability differences. For instance, the higher persistence that we observe for

high-ability teachers may be due to their greater focus on long-term knowledge, or it may be that they are more likely to be assigned students' with a greater capacity to acquire long-term knowledge. While our teacher ability characteristics are only slightly negatively correlated with students' socio-economic and ability characteristics, it is possible that high-ability teachers are assigned more capable students along other, unobserved dimensions. The converse is true for our measures of students' background. Students may have similar aptitude for acquiring long-term knowledge, but may be differentially assigned to teachers who provide lower-quality instruction that emphasizes short-term knowledge rather than long-term knowledge.

In order to further investigate whether differences in persistence are due to instructional factors, we examine whether school-level characteristics predict differences in persistence, independent of teacher and student characteristics. Variation in persistence across schools, rather than across student characteristics within schools, would provide additional evidence that curriculum or instructional quality may be a key factor in differential persistence. For example, school-level differences may reflect school-wide curriculum and instructional practices that influence the degree to which teachers focus on either short- or long-term knowledge. To examine this possibility, we compare ELA value-added persistence across each teacher and student characteristic of interest, for distinct samples of schools that rate either above or below the mean in terms of school-level averages for the same characteristic. In other words, our goal is to identify whether teachers with stronger academic backgrounds have higher persistence even when in schools with relatively few such teachers, and similarly to identify whether poor, black, Hispanic or low-scoring students have lower persistence even when attending schools that are attended by a small portion of these students. In these analyses, we exclude schools where the student or teacher population is extremely homogenous (i.e. >95% or <5%) with regard to our characteristic of interest. We present the results in Table 7.

In samples of schools that are “better-staffed” or that serve fewer poor, black, Hispanic or low-scoring students, we find that both teacher and school characteristics continue to predict differential ELA persistence independent of school characteristics. For example, within a sample of schools where more than a quarter of teachers attended a competitive undergraduate institution, persistence is significantly higher for those students whose teacher attended a competitive institution than for students whose teacher did not, with coefficients of 0.342 versus 0.265, respectively. The same pattern holds true for school-wide averages of student characteristics. For example, within a sample of high achieving schools that had fewer than half of students who were previously low-scoring, teacher value-added persistence for previously low-scoring students is substantially lower than persistence for previously high-scoring students, with coefficients of 0.155 and 0.299, respectively.¹³

In marked contrast to the results for better-staffed schools and for schools serving more advantaged students, schools that have few high-ability teachers or that serve more poor, black, Hispanic, or previously low-achieving students demonstrate low value-added persistence across all individual teacher and student characteristics. For example, in schools where greater than 50 percent of students were previously low scoring, students of all ability levels show very low persistence of learning gains due to their prior teachers’ value-added. The coefficient for persistence within lower-achieving schools is 0.075 for previously low-scoring students, and 0.078 for previously high-scoring students. Similarly, in schools where less than a quarter of the teaching staff attended competitive undergraduate institutions, persistence is low regardless of the type of teacher a student is assigned. In these less-competitively staffed schools, the persistence coefficient for teachers who graduated from competitive institutions is just 0.091, which is identical to the persistence coefficient for teachers from less competitive institutions.

¹³ Differential persistence rates associated with student and teacher characteristics are apparent in more “advantaged” schools across a range of specific cut points used to define our school samples.

School-level poverty, minority, and prior student ability characteristics are highly correlated with each other. However, the negative associations between value-added persistence and schools' student and teacher characteristics represent distinct trends. That is, both the composition of a school's students and the makeup of their teaching staff independently predict differences in persistence rates. As shown in Appendix Table 2, NYC schools that serve a majority of students with above average prior scores and who also hire an above-average proportion of teachers who attended competitive undergraduate institutions have high persistence rates, with a coefficient of 0.353. Schools that fit only one of those two criteria have moderate rates of persistence. Schools that serve primarily students with low prior achievement and who also hire few teachers from competitive institutions have persistence rates that are indistinguishable from zero, with a coefficient of -0.019. Thus, for a sizeable portion of our sample, measures of ELA teachers' value-added effectiveness provide no information at all about their students' longer term academic performance. Both the makeup of the student body, as well as other factors that influence schools' staff composition are important predictors of this trend.

Overall, the variation in persistence that we observe as a function of school-level characteristics suggests that school-wide curricular or instructional factors are influencing students' long-term knowledge gains in ELA. This influence is most dramatically apparent in the very low persistence that we observe in low-achieving schools, even among previously high-achieving students. The persistence of value-added for those high-achieving students is substantially lower than even that of previously low-achieving students who attend high-achieving schools. If differences in student characteristics were the primary driver of differences in value-added persistence, then previously high-scoring students should have retained more long-term knowledge regardless of the quality of instruction that they received. Instead, our results are consistent with a hypothesis that instructional quality is the primary mechanism driving differences in ELA teachers' value-added persistence.

Conclusions and Discussion

Value-added measures of teachers' effectiveness play an increasingly central role in k-12 education policy and practice. They are both practical to implement and directly related to teachers' short-term objectives. In addition, they are correlated with teachers' efficacy at improving students' long-term life outcomes (Chetty et. al, 2011). However, we know relatively little about how and when value added to short-term achievement translates into longer-term gains for students. Focusing too narrowly on measures of short-term productivity may obscure the mechanisms by which teachers create lasting impacts for students or cause us to overlook important heterogeneity in their capacity to do so.

In this paper, we build upon the limited extant research addressing teachers' long-term effects by providing new insights into how teachers influence students' long-term academic performance. In keeping with the prior literature, we distinguish between *short-term* teacher effects that do not persist in following years and *long-term* teacher effects that affect student achievement in the following academic year and potentially beyond. We extend the prior literature by distinguishing long-term effects on *subject-specific* knowledge from long-term effects on *general* knowledge, which benefit students' future performance in both math and ELA. We identify patterns of substantial and systematic heterogeneity in the relationship between teachers', particularly ELA teachers', short- and long-term effects on student knowledge. Our findings highlight both the complexity inherent in assessing teachers' instructional efficacy, and the importance of attending to the persistence of teachers' effects on student learning.

First, we identify substantial crossover effects of ELA teachers on students' long-term math performance, which contrast with the limited crossover effects of math teachers on students' long-term ELA performance. These ELA effects may correspond to the broad relevance of students' reading and writing skills across different subjects and contexts. This hypothesis, if correct, could help to explain an outstanding question suggested by some prior research. On the one hand, ELA teachers' effects on

tested achievement have frequently been observed to be smaller than math teachers' effects (Nye et al., 2004; Rivkin et al., 2004; Chetty et al., 2011). However, because improvements in ELA achievement correlate to greater long-term life outcomes, high value-added ELA teachers on average appear similar to math teachers in terms of their long-term impacts on students (Chetty et. al, 2011). We provide further evidence that teachers' effects on students' ELA knowledge may help students to a greater extent than just their effects on ELA knowledge and skills, through long-term cross-over effects on other subjects. Our findings reinforce the value of investments in student learning in ELA, even if the immediate effects of teachers or other instructional interventions may appear modest in comparison to effects on short-term math achievement. More broadly, these results highlight the potential for important variation in the generalizability of the knowledge that different teachers impart to students.

We also find suggestive evidence that – at least in NYC – instructional factors drive meaningful differences in the extent to which ELA teachers' short-term value added leads to students' longer-term achievement. We see evidence of the importance of instruction in the positive association between teachers' academic ability and their contributions to students' long-term knowledge. Even more compelling, we find that schools that serve more disadvantaged students or that hire fewer of these high-ability teachers have lower value-added persistence in ELA for all of their students. Students, regardless of their prior test performance, who attend schools with many low-performing students demonstrate lower persistence of the learning gains they achieve from having a high value-added teacher. The persistence in low-achieving schools is less than half the rate of that in other schools. These findings provide evidence that instructional quality is a key driver of the variation that we observe in value-added persistence, and that school-level curriculum or instructional norms may foster differences in instructional quality. Unfortunately, we are unable to directly observe the instructional practices of teachers or schools in our sample. However, in light of prior research on educators' responses to high stakes accountability pressures (Corcoran et. al, 2011) one plausible explanation for our findings could

be that schools serving lower performing students systematically prioritize gains in short-term tested achievement in ways that detract from teachers' focus on long-term knowledge generation.

While this study highlights the potential for substantial differences in the persistence of teachers' effects, we can only speculate as to the particular instructional practices that are causing those differences. Though we have access to unique administrative and personnel data about teachers and students, our investigation is exploratory and descriptive, and does not test specific theories about the causal mechanisms underlying variation in value-added persistence. Further research is needed that assesses the effects of instructional practices on persistence. Measures, such as teachers' observed instructional practices or schools' accountability and personnel management practices, would be useful for this work. The rapid expansion of teacher evaluation measures and policies nationwide will likely present many more opportunities for research of this kind.

The variation we observe in the persistence of teacher value-added is likely to be context specific, and thus may not be replicated in all district settings. For instance, our results may stem from specific district-wide instructional practices or incentive systems that do or do not prioritize teachers' short-term knowledge generation. In line with this caveat, we note that at least one other investigation of teacher value-added persistence in North Carolina did not identify comparable differences in persistence rates as a function of students' socio-economic characteristics (Jacob et al., 2010). In addition, variation in the relationship between teachers' short- and long-term effects may be more apparent on some achievement tests than on others. The heterogeneity in persistence that we observe is very pronounced in ELA, but is quite limited in math, and this may be a function of the degree of alignment in explicitly assessed knowledge from one year to the next across these two assessment regimes. Additional research could shed light on the conditions in which teachers' effects on long-term knowledge are likely to vary, as well as the forms of assessment that are most helpful for identifying that variation in long-term knowledge.

Value-added measures allow us to identify teachers' instructional efficacy in a manner that is practical, that is more accurate in terms of predicting future student test-score gains than many alternative measures of teacher performance (Mihaly, McCaffrey, Staiger, and Lockwood, 2013), and that is meaningfully related to teachers' effects on students' life-long outcomes. However, our results indicate that these measures can also obscure important differences in the quality of instruction that teachers and schools provide. In this district, failure to account for either the cross-subject effects of ELA teachers, or for heterogeneity in the persistence of teachers' value-added effects would lead to substantial mis-identification of educators' impacts on students' academic success. In the former case, ELA teachers as a whole might appear to be less influential and important to students' success than they truly are. In the latter, a substantial portion of teachers and schools may in fact be far less effective than they appear to be, when they are assessed in terms of students' acquisition of long-term knowledge and skills, rather than short-term test performance. Failure to accurately identify teacher and school quality is not only inefficient, but could send perverse signals to teachers and school leaders about how best to support students.

References

- Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research*, 74(1), 1-28.
- Buddin, R., & Zamarro, G. (2009). Teacher qualifications and student achievement in urban elementary schools. *Journal of Urban Economics*, 66(2), 103–115.
- Carrell, Scott E. and James E. West. 2010. Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors,. *Journal of Political Economy* 118(3): 409-432.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2011). *The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood* (No. w17699). National Bureau of Economic Research.
- Corcoran, S. P., Jennings, J. L., & Beveridge, A. A. (2011, March). Teacher effectiveness on high-and low-stakes tests. In *Paper originally delivered to the SREE Conference Washington, DC*.
- Goldhaber, D. (2008). Teachers matter, but effective teacher quality policies are elusive. *Handbook of Research in Education Finance and Policy*, 146-165.
- Goldhaber, D. (2002). The mystery of good teaching. *Education Next*, 2(1), 50-55.
- Goldhaber, D., & Theobald, R. (2012) Do Different Value-Added Models Tell Us the Same Things? Carnegie Knowledge Network Briefs. Stanford, CA.
- Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, 95 (7-8), 798–812.
- Jacob, Brian A., Lars Lefgren, and David P. Sims. 2010. The persistence of teacher-induced Learning. *Journal of Human Resources*, 45(4): 915-943.
- Jacob, B. A., & Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *The Quarterly Journal of Economics*, 118(3), 843-877.
- Kane, T. J., & Staiger, D. O. (2008). *Estimating teacher impacts on student achievement: An experimental evaluation* (No. w14607). National Bureau of Economic Research.
- Kinsler, J. (2012). Beyond Levels and Growth Estimating Teacher Value-Added and its Persistence. *Journal of Human Resources*, 47(3), 722-753.
- Konstantopoulos, S., & Chung, V. (2011). The persistence of teacher effects in elementary grades. *American Educational Research Journal*, 48(2), 361-386.

- Lockwood, J. R., McCaffrey, D. F., Mariano, L. T., & Setodji, C. (2007). Bayesian methods for scalable multivariate value-added assessment. *Journal of Educational and Behavioral Statistics*, 32(2), 125-150.
- Marsh, J. A., Springer, M. G., McCaffrey, D. F., Yuan, K., & Epstein, S. (2011). *A Big Apple for Educators: New York City's Experiment with Schoolwide Performance Bonuses: Final Evaluation Report*. Rand Corporation.
- McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of educational and behavioral statistics*, 29(1), 67-101.
- Mihaly, K., McCaffrey, D. F., Staiger, D. O., & Lockwood, J. R. (2013). A composite estimator of effective teaching. *Seattle, WA: Bill & Melinda Gates Foundation*.
- New York City Department of Education (2013). School accountability tools. Retrieved October 23, 2013 from <http://schools.nyc.gov/Accountability/Tools/Report/Default.htm>.
- Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects?. *Educational evaluation and policy analysis*, 26(3), 237-257.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417-458.
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review*, 94(2), 247-252.
- Rockoff, J. E., & Turner, L. J. (2008). *Short run impacts of accountability on school quality* (No. w14564). National Bureau of Economic Research.
- Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *The Quarterly Journal of Economics*, 125(1), 175-214.

Tables

TABLE 1

Summary statistics for students, teachers, and schools in our analytical sample

Variables	New York City
A. Students	
% Free price lunch	63.8
% Reduced price lunch	8.0
% Black	30.2
% Hispanic	36.3
% White	16.4
% Asian	16.5
% Female	51.8
B. Teachers	
% from a competitive undergraduate institution (Barron's rating)	30.0
Average LAST score	252.8 (21.4)
Average SAT verbal score	488.3 (93.2)
Average SAT math score	472.4 (93.3)
Average years of experience in the district	7.3 (6.3)
C. Schools	
Average % of students eligible for free lunch	69.1 (24.0)
Average % of students Black	36.5 (30.1)
Average % of students Hispanic	39.3 (26.2)
Average % of teachers from a competitive undergraduate institution	24.6 (0.150)

Note: Analytical sample consists of students in grades 5 through 8 in school years 2005-06 through 2011-12, for whom prior-year teacher value added data is available. Summary statistics for teachers reflect only teachers for whom value added measures are available. Standard deviations are shown in parentheses for continuous variable measures.

TABLE 2

Correlation matrix of ELA teachers' value added and ability characteristics, classroom characteristics, and school characteristics

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Teacher Characteristics													
(1) Teacher's value added (in ELA)	1.000												
(2) Barron's ranking of teacher's undergraduate competitiveness (1 =most competitive, 4 = least)	-0.012	1.000											
(3) Teacher's NYS Liberal Arts and Sciences Test (LAST) scores	-0.018	-0.349	1.000										
(4) Teacher's SAT verbal scores	-0.041	-0.322	0.687	1.000									
(5) Teacher's SAT math scores	-0.009	-0.346	0.576	0.674	1.000								
Classroom Composition													
(6) Percent of students eligible for free lunch	0.006	0.037	-0.063	-0.068	-0.081	1.000							
(7) Percent of students who are black or Hispanic	0.044	0.021	-0.073	-0.072	-0.104	0.481	1.000						
(8) Percent of students with twice-lagged ELA scores below mean	-0.034	0.060	-0.072	-0.075	-0.083	0.410	0.541	1.000					
School Composition													
(9) Percent of students eligible for free lunch	0.023	0.048	-0.111	-0.108	-0.114	0.527	0.543	0.404	1.000				
(10) Percent of students who are black or Hispanic	0.051	0.013	-0.070	-0.074	-0.108	0.488	0.932	0.503	0.583	1.000			
(11) Percent of students with twice-lagged ELA score below mean	-0.022	0.052	-0.081	-0.087	-0.112	0.518	0.725	0.637	0.616	0.778	1.000		
(12) Percent of teachers from competitive undergraduate schools	0.034	-0.339	0.212	0.180	0.189	-0.138	-0.168	-0.143	-0.173	-0.179	-0.214	1.000	
(13) Percent of teachers with LAST scores in the top third	0.026	-0.223	0.355	0.257	0.227	-0.013	0.028	-0.019	-0.102	0.029	-0.014	0.443	1.000

Note: Correlations are from a teacher-by-year-level sample of ELA teachers between SY 2005-6 and SY 2010-11.

TABLE 3

Estimates for the Persistence of Observed Knowledge, Long Term Knowledge, and Teacher Value Added Effects

	Persistence Estimates		
	Observed Knowledge	Long Term Knowledge	Teacher Value Added Effects
Same-Subject Persistence			
Predicting ELA with ELA			
Coefficient on lagged ELA achievement	0.645 (0.001)	0.946 (0.003)	0.208 (0.019)
First-stage F-statistic	-	73195.7	2005.3
N of students		940357	
Predicting math with math			
Coefficient on lagged math achievement	0.772 (0.001)	0.932 (0.002)	0.189 (0.008)
First-stage F-statistic	-	169533.3	8630.5
N of students		993794	
Cross-Subject Persistence			
Predicting math with ELA			
Coefficient on lagged ELA achievement	0.581 (0.001)	0.639 (0.003)	0.149 (0.018)
First-stage F-statistic	-	73056.6	1856.6
N of students		921283	
Predicting ELA with math			
Coefficient on lagged math achievement	0.612 (0.001)	0.615 (0.002)	0.043 (0.010)
First-stage F-statistic	-	171614.0	8649.1
N of students		971874	

Note: Coefficient for Observed Knowledge from a regression of current achievement on prior achievement. Coefficient for Long Term Knowledge from an instrumental variables (IV) regression of current achievement on prior achievement instrumented with twice-lagged achievement. Coefficient for Teacher Value Added Effects from a regression of current achievement on prior achievement instrumented with teacher value added quality. IV models include controls for current student characteristics and classroom fixed effects.

TABLE 4

Heterogeneity in the persistence of ELA teachers' value added effects on student achievement in ELA and math in the subsequent school year

Subgroup interactions	Same-Subject: ELA on ELA			Cross-Subject: ELA on Math		
	In-group persistence	Out-group persistence	F-test of equal coefficients [p value]	In-group persistence	Out-group persistence	F-test of equal coefficients [p value]
A. Prior teacher characteristics						
Teacher attended a competitive institution	0.274 (0.025)	0.177 (0.020)	0.000***	0.207 (0.024)	0.123 (0.020)	0.000***
Teacher LAST score in top third	0.254 (0.031)	0.177 (0.020)	0.007**	0.198 (0.030)	0.125 (0.020)	0.010*
Teacher SAT math score in top third	0.202 (0.032)	0.172 (0.024)	0.370	0.199 (0.032)	0.115 (0.024)	0.015*
Teacher SAT verbal score in top third	0.242 (0.036)	0.157 (0.024)	0.020*	0.184 (0.035)	0.124 (0.024)	0.107
B. Student characteristics						
Student eligible for free lunch in prior year	0.161 (0.020)	0.262 (0.022)	0.000***	0.109 (0.020)	0.202 (0.020)	0.000***
Student is black or Hispanic	0.150 (0.019)	0.306 (0.025)	0.000***	0.131 (0.020)	0.187 (0.021)	0.003**
Student twice-lagged test score below mean	0.121 (0.021)	0.234 (0.021)	0.000***	0.087 (0.025)	0.178 (0.019)	0.000***
N of students		940357			921283	

Note: Coefficients shown for interaction terms representing teacher value added persistence for in-group and out-group samples across each teacher or student characteristic. Models include interaction terms (not shown) for value added persistence across observations that are missing data for the particular characteristic of interest. ~p<.1 *p < .05, **p < .01, ***p < .001.

TABLE 5

Heterogeneity in the persistence of math teachers' value added effects on student achievement in math and ELA in the subsequent school year

Subgroup interactions	Same-Subject: Math on Math			Cross-Subject: Math on ELA		
	In-group persistence	Out-group persistence	F-test of equal coefficients [p value]	In-group persistence	Out-group persistence	F-test of equal coefficients [p value]
A. Teacher characteristics						
Teacher attended a competitive institution	0.194 (0.013)	0.191 (0.010)	0.835	0.049 (0.016)	0.051 (0.012)	0.877
Teacher LAST score in top third	0.177 (0.015)	0.188 (0.010)	0.484	0.027 (0.019)	0.055 (0.012)	0.151
Teacher SAT math score in top third	0.215 (0.014)	0.173 (0.013)	0.016*	0.040 (0.018)	0.036 (0.015)	0.853
Teacher SAT verbal score in top third	0.214 (0.015)	0.176 (0.013)	0.040*	0.050 (0.020)	0.030 (0.015)	0.378
B. Student characteristics						
Student eligible for free lunch in prior year	0.183 (0.009)	0.201 (0.011)	0.068	0.035 (0.010)	0.055 (0.014)	0.136
Student is black or Hispanic	0.192 (0.009)	0.187 (0.013)	0.682	0.025 (0.010)	0.090 (0.0162)	0.000***
Student twice-lagged test score below mean	0.191 (0.009)	0.180 (0.009)	0.301	0.014 (0.011)	0.066 (0.012)	0.000***
N of students		993794			971874	

Note: Coefficients shown for interaction terms representing teacher value added persistence for in-group and out-group samples. Models include interaction terms (not shown) for the value added persistence of students or teachers who are missing data for the particular characteristic of interest. ~p<.1 *p < .05, **p < .01, ***p < .001.

TABLE 6

Heterogeneity in the persistence of long term knowledge, by subject

Subgroup interactions	In-group persistence	Std. Error	Out-group persistence	Std. Error	F-test of equal coefficients [p value]
ELA					
A. Prior teacher characteristics					
Teacher attended a competitive institution	0.954	0.005	0.944	0.004	0.074
Teacher LAST score in top third	0.955	0.006	0.944	0.004	0.062
Teacher SAT math score in top third	0.950	0.007	0.939	0.005	0.163
Teacher SAT verbal score in top third	0.946	0.007	0.942	0.005	0.563
B. Student characteristics					
Student eligible for free lunch in prior year	0.946	0.004	0.948	0.005	0.604
Student is black or Hispanic	0.939	0.003	0.960	0.006	0.000***
Student twice-lagged test score below mean	0.922	0.004	0.968	0.007	0.000***
N of students			940357		
Math					
A. Teacher characteristics					
Teacher from a competitive undergraduate school	0.930	0.003	0.931	0.002	0.790
Teacher LAST score in top third	0.928	0.003	0.931	0.002	0.333
Teacher SAT math score in top third	0.924	0.004	0.933	0.003	0.055
Teacher SAT verbal score in top third	0.929	0.004	0.931	0.003	0.583
B. Student characteristics					
Student eligible for free lunch in prior year	0.929	0.002	0.936	0.003	0.008**
Student is black or Hispanic	0.926	0.002	0.943	0.003	0.000***
Student twice-lagged test score below mean	0.905	0.004	0.947	0.004	0.000***
N of students			993794		

Note: ~p<.1 *p < .05, **p < .01, ***p < .001.

TABLE 7

ELA Teacher persistence estimates, with interactions of teacher and student characteristics across school-level characteristics

ELA Teacher persistence estimates, with interactions of teacher and student characteristics across school-level characteristics							
School Samples	Teacher or Student Subgroups [interactions]	In-group persistence	Std. Error	Out-group persistence	Std. Error	F-test of equal coefficients [p value]	N of students
A. Above mean ratio of teachers from competitive schools (1) vs below (2)							
(1) >24% competitive	Teacher attended competitive institution (vs others)	0.342	0.032	0.265	0.031	0.007**	444888
(2) <=24% competitive		0.091	0.072	0.091	0.030	0.994	467417
B. Above mean ratio of teachers with high LAST scores (1) vs below (2)							
(1) >23% with high LAST scores	Teacher LAST score in top third (versus others)	0.297	0.044	0.220	0.032	0.027*	432852
(2) <=23% with high LAST scores		0.155	0.075	0.170	0.029	0.834	484205
C. Above mean ratio of free-lunch eligible students (1) vs below (2)							
(1) <66% free-lunch eligible	Student eligible for free lunch (versus others)	0.327	0.042	0.391	0.040	0.017*	338667
(2) >=66% free-lunch eligible		0.158	0.036	0.199	0.041	0.209	318518
D. Above mean ratio of black and hispanic students (1) vs below (2)							
(1) <54% Black or Hispanic	Student is Black or Hispanic (versus others)	0.245	0.042	0.366	0.036	0.000***	324483
(2) >=54% Black or Hispanic		0.178	0.035	0.255	0.049	0.052	311969
E. Below mean ratio of low-scoring students (1) vs above mean (2)							
(1) <50% low scoring students	Student's twice-lagged ELA test score below mean (vs. others)	0.155	0.038	0.299	0.028	0.000***	463236
(2) >=50% low scoring students		0.075	0.029	0.078	0.035	0.917	460066

Note: All models exclude schools whose student or teacher population is overwhelmingly homogenous (i.e >95% or <5%) in terms of the selected student or teacher characteristic. ~p<.1 *p<.05, **p<.01, ***p<.001.

Appendix Table 1

Heterogeneity in ELA teacher persistence estimates when using alternative value added models

Subgroup interactions	A. School Fixed Effects Model			B. Model without student, classroom, or school covariate controls		
	In-group persistence	Out-group persistence	F-test of equal coefficients [p value]	In-group persistence	Out-group persistence	F-test of equal coefficients [p value]
A. Teacher characteristics						
Teacher from a competitive undergraduate	0.293 (0.037)	0.192 (0.026)	0.013*	0.454 (0.015)	0.422 (0.014)	0.001**
Teacher LAST score in top third	0.284 (0.056)	0.177 (0.025)	0.062	0.452 (0.016)	0.411 (0.014)	0.000***
Teacher SAT math score in top third	0.182 (0.053)	0.178 (0.031)	0.955	0.412 (0.017)	0.412 (0.015)	0.994
Teacher SAT verbal score in top third	0.284 (0.059)	0.148 (0.032)	0.028*	0.415 (0.017)	0.409 (0.015)	0.634
B. Student characteristics						
Student eligible for free lunch in prior year	0.190 (0.024)	0.264 (0.030)	0.007**	0.403 (0.014)	0.460 (0.015)	0.000***
Student is black or hispanic	0.184 (0.023)	0.289 (0.033)	0.001***	0.397 (0.013)	0.491 (0.016)	0.000***
Student twice-lagged test score below mean	0.139 (0.027)	0.270 (0.025)	0.000***	0.226 (0.017)	0.438 (0.015)	0.000***
N of students	939804			940357		

Note: School fixed effect model A includes fixed effects in lieu of school-level covariate controls. Model B predicts current student test score with prior-year test score in the same and alternate subject, and attributes remaining variance entirely to individual teachers. Model B includes no other controls for student, classroom, or school characteristics. ~p<.1 *p < .05, **p < .01, ***p < .001.

Appendix Table 2
ELA Teacher persistence estimates across multiple school-level characteristics

Samples	Schools' Student Characteristics	Schools' Staff Characteristics	Persistence estimate	Std. Error	N of students
A. High achieving and "better staffed"	>50% of students with prior scores above the mean	Above mean ratio of teachers (>24%) attended a competitive undergraduate school	0.353	0.036	258857
B. Low achieving but "better staffed"	>50% of students with prior scores above the mean	Above mean ratio of teachers (>24%) attended a competitive undergraduate school	0.147	0.044	190776
C. High achieving but less well staffed	<=50% of students with prior scores above the mean	Below mean ratio of teachers (<=24%) attended a competitive undergraduate school	0.189	0.046	204823
D. Low achieving and less well staffed	<=50% of students with prior scores above the mean	Below mean ratio of teachers (<=24%) attended a competitive undergraduate school	-0.019	0.047	257728

Note: Students' prior scores are twice-lagged ELA test scores. Teachers' undergraduate competitiveness based on Barron's rankings.